

Appendix

In the following, we list the referenced examples for data representation and policy formulation. In the example given in Table 1, a CSV file consisting a column of timestamp-representing values is transformed to an XML structure. The first row of the CSV file is interpreted as containing the identifier of each data item of a row. Each row is considered to be a data entry consisting of possibly multiple data items.

CSV	XML
time	<code><dataset id="h1"></code>
422875	<code> <dataentry id="0"></code>
278522	<code> <time>422875</time></code>
	<code> </dataentry></code>
	<code> <dataentry id="1"></code>
	<code> <time>278522</time></code>
	<code> </dataentry></code>
	<code></dataset></code>

Table 1. Example of data represented in CSV and the corresponding XML transformation.

The scope of the example policy is a dataset with the identifier `h1`. Any data entry d_{ij} of the dataset under consideration that is tagged with `<time>` is pseudonymized in a way that the resulting pseudonym of the data entry d_{ij} includes a utility tag that allows for performing the mathematical operation “addition” on it. This utility option is bound to the role *analyzer*. Any data entry of the dataset `h1` that is tagged with a tag different from `<time>` is omitted and not included in the resulting PsD.

```
<utility_policy id="all">
  <dataset id="h1">
    <dataentry id="all">
      <time>
        <utility>
          <option>
            mathematical operation
          </option>
          <operation>
            addition
          </operation>
          <binding type="role">
            analyzer
          </binding>
        </utility>
      </time>
    </dataentry>
  </dataset>
</utility_policy>
```

```
</dataset>
</policy>
```

For the example described, the pseudonymization tool selects a function that generates Paillier-encrypted ciphertexts as utility tags of the plaintext values. Note that the XML tags of the input file are not considered to be pseudonymized. The resulting PsDs file content is listed in the example of Listing 2.

```
<dataset id="h1">
  <dataentry id="0">
    <time>
      bfYU1qPrasuAddDN0grrBzuLalY
      ...
      Bo3IQw6PhGJqEdvac/oZZ0gZ
    </time>
  </dataentry>
  <dataentry id="1">
    <time>
      23L/n3AzjS1M0Gou3DaA4ak32Z
      ...
      AG9j7KnwHg5QJu4TgjX4Ut89
    </time>
  </dataentry>
</dataset>
```

Listing 2. Pseudonymization of the dataset of Table 1 according to the utility policy of Listing 1.